Artificial Immune System
inspired by Human Immune System

Bait a Trap: Introducing Natural Killer Cells to
Artificial Immune System for Spyware Detection

Complex Adaptive Systems Seminar
Alireza Tashvir
Outline

• Human immune system
• Artificial immune system
• What is Spyware
• Proposed mechanism
• Results
• Future of AIS
Human Immune System

- Its goal
  - Protect our body against foreign invaders and infectious

- Important players
  - Lymphocytes (B-Cells, T-Cells, Natural Killers (NK))

- How it works (work steps)
  - Recognize
  - Attack
  - Memorize

- Why it is important to investigate
  - Powerful information processing capabilities
  - Highly parallel system
  - Use learning, memory and retrieval to solve recognition and classification
How HIS works

- Try to recognize all of the body cells
  - Classified them as self and non-self
    - Non-self are further categorized by type of defensive mechanism
How HIS works

• Encountering antigens use two different approaches:
  • Innate mechanism
    • Non-specific response (defense)
  • Adaptive (acquire) mechanism
    • Specific response by applying hyper mutations of different genes
      • Clonal expansion

• After successful defense
  • Memorizing the antigens and the location they are exposed
Artificial Immune System (AIS)

- Inspired metaphors from HIS
  - Recognition (Self vs Non-self)
  - Feature extraction (by filtering released proteins - cytokine)
  - Diversity (by hyper mutation of different genes)
  - Learning
  - Memory
  - Decentralize controlling (vs nervous system)
Artificial Immune System (AIS)

• Applications

  • Machine learning (e.g. Pattern recognition, Data clustering)
  • Computer security - Malware detection (Virus, Spyware, ...)
  • Fault diagnosis and tolerance
  • Robotics
  • Optimization
  • Scheduling
Spyware (and current popular detection approaches)

• Spyware is designed to make money by stealing users’ privacy or confidential data, rather than harm the computer systems or self-reproduce in the network
  • Hiding their presence (hide its files and registries)
  • Hiding their behaviors (pretend their behaviors are legitimate)

• Regular detection approaches
  • Signature-based
    • Detect known spywares with high degree of accuracy
    • Unable to detect novel ones
  • Behavior-based
    • Can detect partial new spyware with acceptable accuracy
Natural Killers (NKs)

• Powerful weapons to find and kill the latent viruses (Viruses which decreased their activity to escape the attacks performed by HIS)

• They provide some baits to encourage latent viruses to exhibit their activities more obvious (→ can be recognize by immune cells)

• NKs mechanism is introduced by the authors of the article to facilitate the latent spyware detection process
Natural Killers (NKs)

• **Mechanisms**
  • Have different surface receptors
    • These receptors regulate the cell functions by signal transduction

• **Taxonomy of receptors**
  • Inhibitory receptors (IRs)
  • Activating receptors (ARs)
Natural Killers (NKs)

- How NKs decide to kill or leave the binding cell
  - Based on the balance between inhibitory and activating signals
  - If the summation of these signals is negative
    - The cell is left and categorized as a normal cell
  - Otherwise
    - The cell is killed since it is recognized as an infection
      - NK produce proteins (e.g. perforins) to split the target cell cause to expose the virus
Artificial NK model

- Artificial inhibitory signals from processes, files, and registries
- Artificial activating signals from key logging, information collecting and leaking
- Artificial induction cytokine generating artificial user activities in computer systems
Artificial cell representation

• Artificial Natural Killer (NK)
  • NK(NKRs, Fitness, AV, Status, IC)
    • NKRs: Natural Killer Receptors (Inhibitory receptor - IR / Activating receptor - AR)
    • Fitness: Cell adaptability is measured by its fitness (higher fitness -> more adaptable)
    • AV: Activating value (cumulative value increasing by AR signals and decreasing by IR signals)
    • Status: which initially is inactive. If exceeds a threshold, the NK status changes to active
    • IC: each NK can produce a specific type of IC. Since each spyware exhibit different behavior
Artificial cell representation

• Artificial Natural Killer Receptor (NKR)
  • NKR(Type, Ligand, Affinity, Weight)
  • Type: Shows if it is IR or AR
  • Ligand: Shows which source is bound to the receptor (e.g. file and registry expressions of a program)
  • Affinity: determines the value of the perceived signals
  • Weight: for IR, Weight < 0 and for AR, Weight > 0
Recognition and response algorithm

• forall the signals do
  • forall the NKs do
    • set the affinity of all NKR$s$ to 0;
    • get signal $s$;
    • find all NKR$s$ ($mNKR$s) that match $s$;
    • forall the $mNKR$s do
      • $mNKR$.affinity = $s$.value;
    • end
    • calculate the $AV$ of the NK;
    • if $AV \geq TA$ (activating threshold) and NK.status == inactive then
      • NK.status = active;
    • end
  • end
• end

Input: Signals (both inhibitory and activating)
Output: The status of NK (active or inactive)
Recognition and response algorithm

• The AV is computed by:

\[ AV_{after} = AV_{before} + \sum m_{NKR} \text{Affinity} \times \text{Weight} \]

• Once NK activated, generates artificial user activities by IC
  • If spyware does not detect it as a fake activity
    • responds to it
Recognition and response algorithm

- Artificial induction cytokine (IC)
  - Is defined as a series of bogus user activities
  - $IC = \{UA_1, UA_2, ..., UA_n\}$
Recognition and response algorithm

• Artificial IC properties
  • \( IC(\text{\textit{K}_{UA}}, R, C_0, T_I, T_N, f) \)

\( \text{\textit{K}_{UA}} \): kind of IC (e.g. keystroke, file operation, network request)
R: Cycle number (shows induction and non-induction period)
\( C_0 \): initial concentration of IC in the beginning of each cycle
\( T_I \): time span of each induction period
\( T_N \): time span of each non-induction period
\( f \): function of the IC concentration and time in induction period
Evolution process (NK lifecycle)

- Fitness is computed by:
  - $\text{Fitness}_{after} = \text{Fitness}_{before} \times (1 - C_{\text{decay}}) + \sum_{m\text{NKRS}} \text{Affinity} \times |\text{Weight}|$
  - $C_{\text{decay}}$: attenuation coefficient ($0 < C_{\text{decay}} < 1$)
Experiments with real spywares

<table>
<thead>
<tr>
<th>Functions/Features</th>
<th>Actual Spy V3.0</th>
<th>Spybot V1.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keystroke logging</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>File operation logging</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Internet traces logging</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Send logging report</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hiding appearance</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Hiding behavior</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
Four different ICs are introduced

• $IC_{\text{keystroke}}$: Simulate user keystroke
• $IC_{\text{FileOper}}$: Simulate the creation, deletion of files
• $IC_{\text{WebSurf}}$: Simulate opening web pages in a browser
• $IC_{\text{HTTPReq}}$: Generate HTTP requests
Experiment’s results

(a) Actual Spy ($IC_{FileOper}$)

(b) Spybot ($IC_{Keystroke}$)
Experiment’s results

The changes in number of all kinds of ICs in S1 (Actual Spy)
Experiment’s results

The changes in number of all kinds of ICs in S2 (Spybot)
Thank You